

## 人気経路の推薦のための大規模移動軌跡データ処理

姜 仁河<sup>†</sup> 杉山 武至<sup>†</sup> 石川 佳治<sup>† §</sup>

<sup>†</sup> 名古屋大学大学院情報科学研究科 <sup>§</sup> 国立情報学研究所

### 1 はじめに

GPS デバイス、センサ、ソーシャルネットワーク、いろいろな情報源から膨大な軌跡データが発生している。それに対して、大量の軌跡データの処理に関する研究が積極的に行われてきた。その中で、GPS デバイスにより収集された人や車などの移動軌跡データのマイニングに関する研究が特に注目されている。移動軌跡データに人間活動の情報がたくさん含まれているからである。そこから検出された人間活動の習慣やパターンを利用して、我々の日常生活への支援を提供できるとされる。本研究は大規模軌跡データにより人気経路を発見することを焦点に絞って、適切な移動軌跡データの処理手法を提案する。

### 2 関連研究

本研究と関連した研究には主に以下の三つがある。(1) 最短経路・最速経路の発見：ダイクストラ法 [2] はグラフ上で最短経路を発見する一番典型的なアルゴリズムである。[4] では、グラフの辺の通過時間を考慮して、最速経路を時間依存型の最短経路問題として研究した。(2) 軌跡パターンの発見：代表的な軌跡パターンの発見手法は [3] である。(3) 頻繁経路・人気経路の発見：[5] では取り扱われた問題は、ある時間帯の二点間の最頻繁経路である。[1] は二点間の人気経路の問題に対応した初めての研究である。

研究 [1] では生の移動軌跡データをマイニングして、*transfer network* というグラフを構築する。グラフの頂点である *transfer node* は道路の中の重要な場所を代表する。そして、一次のマルコフ連鎖モデルを利用して、各 *transfer node* から別の *transfer node* までの確率を計算する。経路が人気かどうかはその経路が経由した各 *transfer node* から目的地までの確率の積で評価される。そして、二点間の一番人気の経路はそのスコアが一番高い経路となる。

### 3 研究内容

#### 3.1 本研究の基礎

本研究では [1] で提案された *transfer network* を本研究の基礎とする。*transfer network* は確率ベースのモデルに基づいており、明確さがある。人気経路の推薦は、本質的には主観的な問題である。その主観的な

問題をできるだけ客観的に評価するために、確率を経路が人気かどうかの評価尺度とすることに説得力がある

そして、提案された *transfer network* における *Maximum Probability Product* [1] という人気経路の発見のアルゴリズムは、ダイクストラ法と同様に貪欲法で最適解が得られる。ダイクストラ法と同じ時間複雑度  $O(M + N \log N)$  を持つ。そのため、*transfer network* のようなグラフ上でこのような二頂点間の経路の発見問題を効率に解くことができる。

#### 3.2 問題点

##### 3.2.1 推薦された人気経路の適切さ

既存手法によって検出された一番人気がある経路が実際の一番人気の経路ではないケースは多い。原因としては、以下の二点が考えられる。

(1) 人気経路を連結したものは必ずしも人気経路ではない

図 1 において S から A、A から B に行くときの人気経路がそれぞれ S-A, A-B であったとき、S から D に行きたい人に対して、[1] では S-A-B-D という経路を推薦する。しかし、現実の状況では、バイパスを利用した S-C-B-D が好まれる場合も多い。

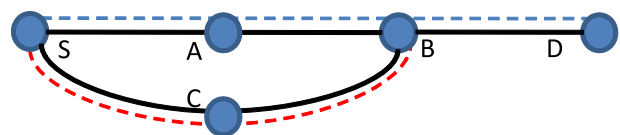


図 1: サブ経路は人気経路ではない例

(2) ノード数が少ない経路を推薦することが多い

[1] は各 *transfer node* の確率値 ( $\in (0, 1)$ ) の積を経路のスコアとして定義したので、*transfer node* を一個多く経由すると、経路全体のスコアが小さくなる。経由ノード数が多い経路は推薦されにくいという問題点がある。

##### 3.2.2 前処理が十分に取られていない

研究 [1] では、*transfer network* 構築などの前処理について効率化を行っていない。そのため、データ量が大きくない場合はうまく対応できるが、データ量が大きくなると、効率が劇的に低下する。

### 3.3 研究課題

本研究は transfer network[1] をベースにして、transfer network 上でユーザにより指定されたスタート地点とゴール地点の二点間の一番人気の経路を発見することを目的とする。具体的に、以上の問題点を考慮して、研究手法 [1] より効率的により正確な人気経路を検出することが本研究の課題となる。

### 4 提案手法

前述した Maximum Probability Product 法 [1] とダイクストラ法 [2] はいずれも一歩ずつグラフを探索していく「one-step」の探索型である。しかし、「one-step」の処理手法はサブ経路も最適解である性質を備える問題にはうまく対応できるが、その性質を持たない人気経路問題において不十分である。そこで、本研究では複数のステップからなる経路をひとまとまりにしたスパン (span) の概念を導入する。[1] と同様、本研究は各 transfer node の確率の積を経路が人気かどうかの基本基準として採用する。一方、実際にその経路を経由した移動軌跡の数も経路の人気の基準の一つとして採用する。それらに基いて、より正確な人気経路を検出するために、以下二つのステップベースとスパンベースのハイブリッド型の処理手法を提案する。

#### 4.1 修正処理の追加

アイデアとしては、欲張り法のアルゴリズムで最大の確率の経路を探索する途中で、適当なタイミングでスタート頂点から今訪問している頂点までの部分経路が最適解であるかどうかをチェックする。チェックの基準としては、実際にこの部分経路を経由した移動軌跡の数を利用する。部分経路  $SR_1$  と  $SR_2$  について、それぞれの確率の値を  $P_1$  と  $P_2$  で、実際に経由したパスの数を  $N_1$  と  $N_2$  で表す。閾値は  $\delta$  と  $\theta$  になる。チェックの条件  $0 < (P_2 - P_1) < \delta$  かつ  $(N_1 - N_2) > \theta$  に合致した場合、 $SR_1$  を最適な部分経路として、 $SR_1$  から続いて次の経路を探索していく。この手法は「one-step」と「one-span」を併用する。これにより部分経路を適切に修正できると考える。

#### 4.2 部分人気経路のサマリを作成

アイデアとしては、前処理により、前述した拡張基準で生の軌跡データからスパンに基づく部分人気経路を検出して、図 2 に示すように、各頂点で別の頂点までの部分人気経路を格納する。すると、transfer network 上で頂点  $v$  を訪問するとき、ゴール地点  $D$  により部分人気経路  $SR$  を得て、直接に  $SR$  から次の人気経路を探索していくことが可能となる。具体的に、軌跡パターンの発見の処理手法 [4] を活用して、適切な長さの軌跡パターンを発見して、それを部分人気経路とする。この手法は前処理の段階で「one-span」のサマリを作成して、「one-step」と「one-span」のハイ

ブリッドの探索型を採用する。それで部分経路を適切に修正できるだけでなく、経路探索の効率性も向上できると考える。

$V_i$	$V_j$	頂点V	部分人気経路SR
		$V_1$	$V_3 \rightarrow V_6 \rightarrow V_7$
$V_1$		$V_8$	$V_3 \rightarrow V_5$
$V_{10}$		$V_9$	NULL
...	...	$V_n$	$V_4 \rightarrow V_{12} \rightarrow V_{n-2}$

図 2: 各ノードで管理されるサマリ表

### 5 まとめ

本研究は二点間の人気経路の発見の問題に対して、「one-step」と複数ステップからなる「one-span」のハイブリッド型の修正技術と前処理技術を提案した。重要な点としては、以下の二点がある。

(1) 既存手法 [1] に比べて、より正確な人気経路を検出することが可能である。

(2) 前処理技術を利用することで、処理効率を向上させる。大量の軌跡データにも対応可能とする。

### 6 今後の課題

本稿では、移動履歴データをもとにユーザに人気の経路を推薦するアプローチについて述べた。既存の研究 [1] における「人気経路の連結により人気経路が得られる」という仮定は必ずしも正しくないことから、スパンの概念を導入し、より現実的な人気経路の推薦を目指した。今後の課題としてはアルゴリズムの詳細化と実験に基づく評価が挙げられる。

### 謝辞

本研究の経費の一部は科学研究費 (25280039) および地球環境情報統融合プログラム (DIAS-P) による。

### 参考文献

- [1] Zaiben Chen, Hong Tao Shen, and Xiaofang Zhou. Discovering popular routes from trajectories. In *ICDE*, 2010.
- [2] Edsger W. Dijkstra. A note on two problems in connection with graphs. *Numerische Math*, 1, 1959.
- [3] Fosca Giannotti, Mirco Nanni, Fabio Pinelli, and Dino Pedreschi. Trajectory pattern mining. In *SIGKDD*, 2007.
- [4] Evangelos Kanoulas, Yang Du, Tian Xia, and Donghui Zhang. Finding fastest paths on a road network with speed patterns. In *ICDE*, 2006.
- [5] Wuman Luo, Haoyu Tan, Lei Chen, and Lionel M. Ni. Finding time period-based most frequent path in big trajectory data. In *SIGMOD*, 2013.